DOCUMENT RESUME

ED 182 707 CS 005 261

AUTHOR Hanson, Marjorie: And Others
TITLE Statistical Models and Practical Significance in Reading Research.
PUB DATE 79
NOTE 16p.: Paper presented at the Annual Meeting of the National Reading Conference (29th, San Antonio, TX, November 29-December 1, 1979)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Reading Research: *Research Methodology: *Research Needs: *Research Problems: *Statistical Analysis: *Tests of Significance

ABSTRACT

Reading researchers should be encouraged to include assessments of practical significance in all their journal articles, reviews of research, technical reports, and dissertations. The tendency has been to base the integration of data entirely on the results of tests for the significance of differences, overlooking the assessment of practical significance. A review of the research studies reported in the 1977-79 yearbooks of the National Reading Conference shows that few discussions included assessments about the strength of relationships or practical significance. This failure to assess the strength of relationships not only limits the information value of individual studies or experiments, but also constrains opportunities to synthesize empirical findings and to advance theories about learning to read and reading instruction. When "highly significant" ANOVA test results are accompanied by "extremely low" values of practical significance, researchers should realize that they "know something for sure about very little." When high values of practical significance repeatedly appear for the same instructional variables, the potential value of these variables in the formulation of reading theory should be recognized and exploited. (RL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED182707

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Marjorie Hanson, Research Associate
Department of EdAd

William H. Rupley, Associate Professor
Department of EdCI (to receive all correspondence)

James F. McNamara, Professor
Department of IdEd

Texas A&M University
College of Education
College Station, Texas 77843

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Marjorie Hanson, William H. Rupley, James F. McNamara
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Statistical Models and Practical Significance in Reading Research*

Marjorie Hanson
Texas A&M University

William H. Rupley
James F. McNamara
Texas A&M University

Two kinds of statistical tests of particular interest to researchers are those that test for the significance of relationships and those that test for the significance of differences. To determine whether there are significant differences among two or more groups on a particular criterion variable outcome measure of interest, researchers frequently employ a univariate analysis of variance. This statistical model and its unique set of decision rules allows one to specify directly the statistical significance associated with the test of an experimental hypothesis of interest. Practical significance, on the other hand, depends on an accurate estimate of the strength of a statistical association. The practical significance assessment usually follows the design employed in tests for the significance of relationships, and often begins by asking "How much of the variance in a criterion measure can be accounted for by a prediction measure?"

*Preparation of this article was supported in part by the Reading/Language Research Laboratory and in part by the Instructional Research Laboratory, Department of EdCI, Texas A&M University, College Station, TX.

CS005261

Both statistical and practical significance are based on information assembled for the traditional F-tests employed in ANOVA designs (NcNamara, 1979). However, there has been a tendency among educational researchers to base the integration of data entirely on the results of tests for the significance of differences (McNamara, 1978; McNamara and Gill, 1979). Overlooking practical significance; that is, the failure to assess the strength of relationships, not only limits the information value of individual studies or experiments, but also constrains opportunities to integrate or synthesize empirical findings as a means to advance our theories regarding learning and instruction. The latter includes current efforts in secondary analysis (Burnstein, 1978) and meta analysis (Glass, 1976) as well as the less frequently utilized statistical synthesis methods based on cluster sampling strategies (Jackson, 1978).

The intents of this inquiry were to classify the statistical models employed in research studies reported in the 26th, 27th and 28th Yearbooks of the National Reading Conference, and to assess the practical significance for a particular subset of ANOVA findings.

Content Analysis

In an attempt to ascertain whether or not the reading research reported in the three most recent NRC Yearbooks addressed the practical significance of results, a content analysis of each reported investigation was conducted. Table 1 presents the initial stage of this analysis, that is a summary of statistical models employed by researchers for each yearbook of interest.

Statistical Models

Table 1 clearly illustrates that some type of analysis of variance

framework was used by a majority of investigators. When t testing applications are added to the ANOVA designs, it can further be seen that almost all significance testing employs parametric statistical models. For example, only 5 of the 45 statistical models identified for the 1977 Yearbook were nonparametric frameworks.

Insert Table 1

The Table 1 entries are comprehensive but not complete, since a few studies were impossible to classify primarily because discussion of analyses performed was inconsistent with standard reporting procedures. Specifically, two studies in 1977 failed to provide adequate information on supplementary parametric and nonparametric analyses and a third study (1979) reported significance levels but failed to identify the specific statistical model that produced the test results. A fourth article (1979), which could not be classified, was a well-designed synthesis of current research in a specific domain of reading. Finally, correlations that were utilized explicitly for analyses of test reliability and consistency were not entered in Table 1.

While this itemized list of statistical models is interesting, the actual focus of this content analysis was to determine the amount of attention given to discussion of the strength of relationships among factors that are found to be significant. Although ANOVA frameworks

can also be formulated as multiple linear regression models (Cohen, 1968) this contemporary approach to analysis was almost never employed during this three year period. If this option would have been incorporated in reading research designs, consideration of the practical significance question - the amount of criterion variance accounted for by the ANOVA treatment variables - would have been an integral part of the model output.

Practical Significance

Few discussions of strength of relationship, or responses to the practical significance question can be found in research reported in the three most recent NRC Yearbooks. Among the majority of studies that utilized analysis of variance designs, only four (1 in 1977, 1 in 1978, and 2 in 1979) discussed the variance accounted for by the independent variables in their designs. One nonparametric study in 1977 reported a single Cramer's V to reflect the relationship between two variables used in a Chi-square test. A study utilizing factor analysis in 1977 also reported the discriminant analysis as a test for the importance of fingings derived from the factor analytic model. Although the frequently reported Pearson product moment correlation lends itself to easy consideration of strength of relationships, the r-squared value which readily serves this purpose was seldom calculated or discussed in correlational studies.

Multiple regression models contain considerations of strength of relationships as an integral part of analysis and, consequently, if

any statistical results are reported from regression analysis, R^2 coefficients are included. This was true of the multiple regression models reported in the yearbooks; all studies utilizing multiple regression that reported results also included statistical measures of the strength of relationship (R^2).

Conversely, the analysis of variance models that dominated the empirical studies and also formed the focus of this content analysis rarely took the opportunity to report a related measure of practical significance. However, such a measure can easily be calculated from the statistics conventionally provided in analysis of variance studies. To illustrate, η^2 (Greek eta, squared) one widely accepted measure of practical significance is calculated as follows,

$$\eta^2 = \frac{F\,[df(n)]}{F\,[df(n)] + df(d)}$$

where F is the reported valve for the relevant ANOVA test of significance; df(n) is the numerator degrees of freedom for the F-statistic; df(d) is the denominator degrees of freedom for the F-statistic; and η^2 is an indicator of practical significance. Eta-squared can also be viewed as the ratio of the appropriate sum of squares for treatments to the total sum of squares for a given ANOVA model (McNeil, Kelly and McNeil, 1975). Alternate practical significance indices that have been introduced in behavioral science literature include those advanced by Golding (1975), Halderson and Glasnap (1972) and McNamara (1978, 1979).

Since accepted reporting of analysis of variance design normally includes the data requirements specified in the η^2 formula offered above, we were able to calculate this practical significance index for most single factor univariate analyses reported in the three yearbooks.[1] The results of these straight forward calculations are recorded in Table 2. In view of the noticeably high η^2 values indicated in the table, it is somewhat puzzling that researchers in reading are not more open about reporting them, as in most cases they add substantially to the strength of their findings.

Insert Table 2 here

Based on the results of this intensive content analysis, the authors offer the following for careful consideration.

(1) The η^2 statistic or some other measure of practical significance would be useful in directing reading researchers' attention to those variables that warrant further study in subsequent investigations.

(2) Too frequently insufficient data are provided in research reports in the NRC Yearbook to enable an assessment of the practical significance of results by consumers of research studies. Descriptions of designs were often imbedded in the text and not sufficiently clear to enable the reader to reconstruct the complete ANOVA table.

(3) Further, even in those studies where we were able to calculate the η^2 statistics, there was often some question as to the accuracy of the degrees of freedom. In these cases, if the description of the sample and of the treatment variables had been comprehensive, the model could have been reconstructed and the correct degrees of freedom ascertained.

(4) A need for more accurate descriptions of sampling plans was evident since sample selection is likely to have a decided influence on both tests of differences and tests of relationships. Further, discussion of generalizability were notably absent from the empirical studies that were reviewed in the yearbooks. Such a discussion is especially vital in view of the extremely small and non-random samples which were typically utilized.

(5) In several of the ANOVA studies a series of univariate tests were performed utilizing the same independent variables to assess significant differences for several distinct criterion variables without any reference to the possibility that one or more of the resulting individual F test differences might be due merely to chance since the set of F tests are not statistically independent. In such cases a single multivariate design would have been more appropriate.

(6) It can be noted, also, that nonparametric techniques are seldom used, although the characteristics of the researcher-developed instruments utilized in some of the studies suggests

that the measurements would fail to meet at least the assumption of continuity in the dependent variable.

Implications for Reading Researchers

While our treatment of practical significance in reading research focused primarily on the use of η^2 to aid in the interpretation of the importance and magnitude of differences identified in analysis of variance tests of significance, it is not implied that attention to practical significance represents a concern just recently entered into the educational literature. Practical significance in educational research was initially addressed in the mid 1960s (Schutz, 1966). However, the content analysis does point out that in 1977 through 1979--eleven to thirteen years later--attention to the practical significance has not yet found its way into the reading research published in the NRC Yearbooks.

Reporting Research

Most statisticians agree that η^2 (or a comparable index of practical significance) communicates a great deal more information than the traditional F value. It takes into account sample size and thus provides a more accurate representation of the degree of prediction. Furthermore, since η^2 is calculated directly from ANOVA results, it creates no difficulty for the investigator.

Given the fact that considerable interest is presently being given to meta analysis and secondary analysis of educational research,

reading researchers' reporting of practical significance needs to be given paramount consideration. Although one cannot accurately specify how often the research reported in any given reading journal is consulted or cited as a reference for various decision making situations, authors should report an index of practical significance. By putting their readers on notice as to their position on practical significance requirements, researchers no longer restrict the users of such research to merely referencing or counting the number of statistically significant differences on criterion variables of interest.

It is our contention that reading researchers should report η^2 (or a comparable index) in journal articles, reviews of research, technical reports and dissertations. Also, practical significance values should be in the forefront when any reading researcher finishes the results section of his or her study and enters the section on discussion of findings.

Training Researchers

To demonstrate that the η^2 value communicates a good deal of information, and thus provides reading researchers more accurate measures of the relative reduction of uncertainty that can be directly linked with a particular research inquiry, professors who train future researchers or program evaluators might initiate one or more of the following alternative activities.

Alternative One. In seminars devoted to reading research design, an interesting project might be to conduct a practical significance

analysis on all recently completed doctoral dissertations. When classification errors (low practical significant values for highly significant difference tests) are encountered, students could be asked to prepare revised interpretations and discussions of dissertation results. Comparisons of original and revised versions provide excellent means to illustrate the consequences of viewing tests of significance on an experiment as the ultimate objective.

Alternative Two. In seminars devoted to theory development or knowledge synthesis, professors should encourage students to compute and include η^2 values (or their equivalent) for each research study reviewed. In the past, probability values (i.e., $p \leq 0.01$) have generally been used to support theories. The intent here would be to suggest that an index of practical significance rather than probability values should be used to indicate the satisfaction students have with their theoretical frameworks.

Alternative Three. If professors are able to locate current operational studies where the probability values associated with mean difference ANOVA tests were "highly significant" but the η^2 values were "extremely low," they might use a seminar group to conduct an organizational analysis (case study) to determine how administrators and other policy-making groups in the organization actually use their "significant differences" tests results in subsequent decisions (i.e., reporting results to the public, creating new priority issues, preparing press releases, changing teacher assignments, or implementing new instructional designs). Analysis of an actual organizational example

and the real consequences that are uncovered in that setting may be far more convincing to graduate students whose immediate professional interests center primarily on practice rather than theory development and verification.

Summary

The implications suggested above are intended to give reading researchers the confidence to note the following. First, when "highly significant" ANOVA test results are accompanied by "extremely low" eta squared values, one would do well to follow the advice of McNeil and his colleagues (1975) who interpret this situation as "knowing something for sure about very little." Second, when high η^2 values are continually uncovered on the same instructional variables, the field should be encouraged to recognize their potential value for the development of theory in reading research.

Hopefully, awareness of these facts will encourage the reporting of practical significance in future reading research investigations.

NOTES

1. Due to lack of clarity in the discussion of factorial and multivariate designs as to the orthogonality or proportionality of the designs, we were unsure of the model or weighting procedures that had been utilized in these analyses. Therefore, these higher level factorial tests were not included in our calculation of practical significance.

REFERENCES

Burnstein, L. J. Secondary analysis: An important resource for educational research and evaluation. Educational Researcher, 1978, 5, 9-12.

Cohen, J. Multiple regression as a general data-analytic system. Psychological Bulletin, 1968, 70, 426-443.

Glass, G. V. Primary, secondary and meta analysis of research. Educational Researcher, 1976, 10, 1-4.

Golding, S. L. Flies in the ointment: Methodological problems in the analysis of the percent of variance due to persons and situations. Psychological Bulletin, 1975, 82, 278-288.

Halderson, J. S. and Glasnap, D. R. Generalized rules for calculating the magnitude of an effect in factorial and repeated measures ANOVA designs. American Educational Research Journal, 1972, 9, 301-310.

Jackson, G. B. Methods for reviewing and integrating research in the social sciences. Final report, National Science Foundation Grant # DIS 76-20398. Washington, D.C.: Social Research Group, George Washington University, April 1978.

McNamara, J. F. Practical significance and statistical models. Educational Administration Quarterly, 1978, 14, 1, 31-50.

McNamara, J. F. and Gill, D. R. Practical significance in vocational education research. Journal of Vocational Education Research, 1979, 3, 1, 20-38.

McNamara, J. F. Practical significance and mathematical models in administrative research. In G. L. Immegart and W. L. Byrd (Eds.). Problem-finding in educational administration: Trends in theory and research. Toronto: Lexington Books, 1979.

McNeil, K. A., Kelly, F. J. and McNeil, J. T. Testing Research Hypotheses Using Multiple Linear Regression. Carbondale, Ill.: Southern Illinois University Press, 1975.

Schutz, R. E. Research in schools. Educational Researcher, 1966, 5, 151-158.

Table 1

Frequency of Statistical Models Employed in the 1977, 1978, and 1979 Yearbooks of the National Reading Conference

Types of Statistical Models	Number Reported*			
	1977	1978	1979	Total
t-tests (both independent and correlated)	7	10	5	22
Analysis of Variance	21	31	20	72
single level univariate	10	4	7	21
factorial univariate (including analysis of covariance)	11	23	12	46
multivariate	0	4	1	5
Non-parametric	5	3	3	11
Multiple Regression	2	3	5	10
Factor Analysis	1	1	2	4
Pearson Product r	7	3	7	17
Total empirical studies	33	33	33	99
Total studies	57	49	55	161

*The number of models applied does not equal the number of studies with empirical tests; several researchers utilized more than one type of test.

Table 2

Indicators of Practical Significance for Single Level Univariate Models* Reporting Statistical Significance

Eta Squared Ranges		Yearbooks 77	78	79	77-79
.0100	.0199		1		1
.0200	.0299	1			1
.0300	.0399		1		1
.0400	.0499	1	1	1	3
.0500	.0599	3			3
.0600	.0699	1	2		3
.0700	.0799	2	1	1	4
.0800	.0899	3			3
.0900	.0999	2	3		5
.1000	.1999	8	6	2	16
.2000	.2999	9	4	10	23
.3000	.3999	6	3	4	13
.4000	.4999	1		2	3
.5000	.5999	1	1	2	4
.6000	.6999		1	1	2
.7000	.7999		1	1	2
Number of tests		38	25	24	88
Number of studies (with sufficient information)		10	8	8	27
Number of studies (with insufficient information)		4	2	1	7

*T-tests of significance are assumed to be independent and thus equivalent to a single-level dichotomous F in the calculation of these eta^{2}s. One series of t-tests (1979) which are clearly identified as correlated is omitted from the calculations as well as post-hoc tests of significance in ANOVA designs.